Symptoms of Depression in Adolescents

Examining the psychometric properties of a Norwegian translation of the Short Mood and Feelings Questionnaire

Ken-Stian Solbakken Olsen

Master of Philosophy in Developmental Psychology

Department of Psychology
UNIVERSITY OF OSLO

May 2015

2015

“Symptoms of Depression in Adolescents – Examining the psychometric properties of a Norwegian translation of the Short Mood and Feelings Questionnaire.”

Ken-Stian Solbakken Olsen

http://www.duo.uio

Trykk: Reprosentralen, Universitetet i Oslo

Acknowledgements

Author: Ken-Stian Solbakken Olsen.

Supervisors: Evalill Bølstad Karevold & Dag-Erik Eilertsen (University of Oslo).

Title: Symptoms of Depression in Adolescents – Examining the psychometric properties of a Norwegian translation of the Short Mood and Feelings Questionnaire.

Among the most important of qualities in both a scientist and a person are having an inquisitive mind and a thirst for knowledge. Science is the greatest tool ever devised and to be able to contribute to the field of psychology is truly inspiring. Working under, at times, massive pressure and strenuous deadlines, while still trying to be both a friend and a boyfriend has been an unparalleled learning experience and, not surprisingly, the subject about which I have learned the most is myself. I sincerely hope that my master's thesis will be as valuable to the field of psychology as the experience itself has been for me.

In science, one is standing on the shoulders of giants; I would therefore like to thank all contributors to the field of psychology – by building on previous discoveries, one can come closer to the truth. I would also like to thank all contributors to the TOPP study and its informants, as well as the Norwegian Institute of Public Health and the University of Oslo for enabling me to conduct my own research. My supervisors, Evalill and Dag-Erik, have been invaluable, and they have made me challenge my own limits. For this, I thank you. Malene and Martin have read my thesis along the way, provided valuable pointers and helped me shape my work to a higher level. I also owe my friends and family a debt of gratitude. Last, but certainly not least, I have to thank Trine for her patience, tenacity, guidance, support and comfort during the last two years. I know I could not have finished this thesis without you.

The author and the supervisor Evalill Karevold developed the idea for the study. The author and both supervisors designed the analyses, which are based on already collected data from the TOPP study. Most analyses were performed by the author.

Ken-Stian S. Olsen
Oslo, May 2015

Abstract

In order to provide proper preventive care or implement appropriate interventions, one needs valid, reliable and thoroughly vetted tools for mapping and identifying at-risk or affected children and adolescents. These tools need to possess satisfactory psychometric abilities in order to be used as tools for identification, screening or evaluation. The current study examined the psychometric properties of the Short Mood and Feelings Questionnaire (SMFQ), and the depressive symptomatology in a community-based sample of 12-13 year old adolescents and their mothers, based on data from the TOPP-study. Findings from the factor analysis supported a unidimensional interpretation of the SMFQ, thereby replicating findings across culture and language. The SMFQ had good reliability, appeared to have adequate content validity, and the parent-child agreement was moderate. Children scored significantly higher than mothers, and girls scored significantly higher than boys. The ratio of girls to boys operationalised as depressed was 4 to 3. Boys who had single mothers, and girls living in families with lower income and an unsatisfactory economy had more depressive symptoms. The use of SMFQ sum-scores as a continuous measure of severity of depressive symptoms can be supported.

Keywords: Depression, adolescents, SMFQ, validation, psychometrics, epidemiology, socioeconomic status.

Sammendrag

For å kunne gi god forebyggende behandling eller iverksette hensiktsmessige tiltak, må man ha valide, pålitelige og grundig studerte verktøy for å kartlegge og identifisere utsatte eller berørte barn og unge. Disse verktøyene må ha tilfredsstillende psykometriske egenskaper for å bli brukt som verktøy for identifisering, screening eller evaluering. Denne studien undersøkte de psykometriske egenskapene til SMFQ, i tillegg til depressiv symptomatologi i et populasjonsbasert utvalg av 12-13 år gamle ungdommer og deres foreldre, basert på data fra TOPP-studien. Faktoranalysen støtter en undimensjonell tolkning av SMFQ, hvilket repliserer funn på tvers av kulturer og språk. SMFQ hadde god reliabilitet, adekvat innholdsvaliditet, og foreldre-barn enigheten var moderat. Barn skåret signifikant høyere enn mødre, og jenter skåret signifikant høyere enn gutter. Forholdet mellom jenter og gutter operasjonalisert med depresjon var 4 til 3. Gutter med single mødre, og jenter i familier med lavere inntekt og utilfredsstillende økonomi hadde flere depressive symptomer enn gjennomsnittet. Resultatene støtter bruken av sum-skårer på SMFQ som et helhetlig mål på alvorlighetsgrad av depressiv symptomatologi.

Nøkkelord: Depresjon, ungdom, SMFQ, validering, psykometri, epidemiologi, sosioøkonomisk status.

Table of Contents

Introduction

"Depression is the inability to construct a future". (May, 1989)

Depression can be a crippling disorder and can lead to broken lives and broken families. Depression is a major concern for public health and causes great societal costs every year (Sanne, Dahl & Tell, 2001). Mental disorders cost Norway more than any other group of afflictions, and depression accounts for about half of all the costs for mental disorders (Holte, 2012). Depression often occurs for the first time in childhood or adolescence and has a high degree of recurrence in adulthood (Sund, Bjelland, Holgersen, Israel & Plessen, 2012). Thus, it can be argued that depression is one of the most important afflictions where proper prevention, early identification and good treatment would be of great significance for both individuals and the society.

Mental disorders are common and often of early onset, the prevalence of which only increases through development (Heiervang et al., 2007). Depression in adolescence and early adulthood is often recurring, and can lead to adverse outcomes such as anxiety, suicidal tendencies and welfare dependence (Fergusson, Boden, & Horwood, 2007). Symptoms of depression in adolescence can strongly predict depression in adults (Pine, Cohen, Cohen & Brook, 1999). In a study of depressive symptomatology in Norwegian adolescents, it was theorized that most of adolescents who have symptoms of depression have problems related to everyday functioning (Lundervold, Breivik, Posserud, Stormark & Hysing, 2013). If we can identify and provide care to children and adolescents suffering from mental distress as early as possible, it would be easier to prevent mental disorders from occurring later in life (Mykletun, Knudsen & Mathiesen, 2009). In a study of the prevalence and characteristics of depression in Norway, the authors conclude that there is a need for improved assessments of young individuals (Sund, Larsson & Wichstrøm, 2011).

Identification, screening and assessment of children and adolescents who are at risk of developing, or already have developed, depression, is very important. In order to provide proper preventive care or implement appropriate interventions, one needs valid, reliable and thoroughly vetted tools for mapping and identifying at-risk or affected children and adolescents. These tools need to possess satisfactory psychometric abilities in order to be used as tools for identification, screening or evaluation. The development and validation of instruments is a continuous and ongoing process (e.g. Hunsley & Meyer, 2003; Messick, 1995; Turner, Joinson, Peters, Wiles & Lewis, 2014), where the goal is to incrementally improve the knowledge about an instrument's psychometric abilities.

The TOPP study (Tracking Opportunities and Problems in Childhood and Adolescence) is a longitudinal study that started in 1993 and that gathered questionnaire data from children, adolescents and their families at eight different time points. The study examines developmental pathways to good mental health and mental disorders in children and adolescents, as well as mental health, relationships and work related issues among parents. One of the subscales used in the TOPP study is the Short Mood and Feelings Questionnaire.

The Short Mood and Feelings Questionnaire (SMFQ) is an instrument designed for the quick evaluation of depressive symptomatology in children and adolescents (Angold, Costello, Messer & Pickles, 1995; Messer, Angold, Costello & Loeber, 1995), based on a longer version called the Mood and Feelings Questionnaire (MFQ; Angold, Costello, Pickles, Winder & Silva, 1987). There are questionnaires for both child self-report and parental report. The SMFQ is a tool for studying depressive symptoms in adolescence, it is an appropriate depression instrument for use in prospective studies from childhood through to late adolescence, and it is able to distinguish between depressed and non-depressed subjects in the general population (Angold et al., 1995; Turner et al., 2014). In a review of research studies incorporating either the SMFQ or the MFQ, it is concluded that only the MFQ, the long version, has been psychometrically examined in Norway (Richter & Sund, 2013). The internal structure of the SMFQ has been examined, but not in a sample of 12-13 year adolescents and their parents. The SMFQ is widely used, but most studies examining the validity of the SMFQ have focused on selected or clinical samples in childhood, and few have examined the SMFQ in community samples (Turner et al., 2014).

The current study focuses on a community-based sample in Norway, comprising of adolescents in their early teens and their mothers, using data from the TOPP study. The primary aim of the current study is to examine the psychometric properties of the SMFQ and validating it in this sample. The secondary aim of the current study is to see if the use of SMFQ sum-scores as a continuous measure of severity of depressive symptoms can be supported, through which differences in depressive symptomology and parent-child agreement can be examined. The relationship between socioeconomic status and depressive symptoms are also of interest. Thus, the overarching goal of the current study is to improve the knowledge about the SMFQ's psychometric abilities. An examination and validation of the SMFQ for use in a community-based sample of Norwegian adolescents in their early teens, as well as their mothers, would be of interest for further research and epidemiology, as there is always demand for demonstrably valid instruments.

Theoretical and Empirical Background

Psychometrics & Validation

Validity has been defined as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores" (Messick, 1989, p. 13). In other words, it is not the questionnaire that is valid in itself – validity speaks to whether the actions we take based on questionnaire scores can be supported empirically and theoretically. If we find that the SMFQ accurately measures symptoms of depression in adolescents, we can use the scores on the SMFQ to say something about depressive symptomatology in adolescents.

Internal structure. The primary concern in validation is assessing construct validity through examining the degree to which an instrument represents the underlying construct (Messick, 1995). Depression cannot be observed directly. Therefore, one has to use instruments to examine the underlying construct, and validity refers to which degree the instrument and one's theoretical understanding of the underlying concept coincide (Turner et al., 2014). An important part of examining construct validity is assessing the structural validity of a questionnaire or test (Loevinger, 1957). The internal structure of a questionnaire should reflect what is known about the internal structure of the construct (Messick, 1989). Depression is generally thought to be unidimensional (Aggen, Neale, & Kendler, 2005). The unidimensionality of a measure of depressive symptomatology addresses the ability to discriminate between distinct groups and therefore speaks to the validity of the measure (Costello & Angold, 1988). Since depressive symptoms generally increase with age, it is important to examine whether patterns of SMFQ symptom covariation vary with age (Weiss & Garber, 2003; Messer et al., 1995). Generalisations regarding developmental changes in depression are partly dependent on the psychometric validity and internal structure of instruments, i.e. if one is to say that there are changes in self-reported depressive symptomatology during childhood and adolescence, one needs a unidimensional depression measure (Messer et al., 1995).

If we accept that item endorsement on the SMFQ is the same as symptomatology of depression, then patterns of symptom covariation can be described as the factor structure of the SMFQ, and the internal structure of the SMFQ can be examined through a factor analysis (John & Benet-Martinez, 2000; Messer et al., 1995). Empirical evidence suggests that the SMFQ measures a unidimensional construct, in that it seems to not measure different types of constructs, and rather seems to measure depression as a single latent variable (e.g. Angold et

al., 1995; Angold, Erkanli, Silberg, Eaves & Costello, 2002; Costello, Benjamin, Angold & Silver, 1991; Messer et al., 1995; Sharp, Goodyer & Croudace, 2006). The essential unidimensionality of the SMFQ was confirmed in a cross-sectional study in Western Norway where the SMFQ was administered to over 9000 adolescents, aged 16-19 (Lundervold et al., 2013). There is insufficient information regarding the internal structure of the SMFQ in a sample of 12-13 year old adolescents and their parents in Norway.

Reliability. Generally, judgements of validity are focused on whether an instrument measures what it is meant to measure, whereas judgements of reliability are focused on the measurement's accuracy. Validity and reliability are closely related, in that an instrument cannot be considered to be valid unless it is reliable, but an instrument can be reliable but not valid (Tavakol & Dennick, 2011). Reliability is used to estimate the consistency or accuracy of results across items within a test (Cortina, 1993). Reliability estimates can be given as Cronbach's alpha (α), represented as a number between 0 and 1. Cronbach's alpha can be interpreted as the percentage of the observed individual differences attributable to true variance (Cronbach, 1951; Cronbach & Shavelson, 2004), as the mean of all possible split-half coefficients (Cortina, 1993), or as an estimate of the proportion of variability in the measure, attributable to the true score (Trochim, 2006). Cronbach's alpha is a function of two parameters: The interrelatedness of items in a scale and the number of items. In other words, scale length is important when interpreting alpha. If a scale or questionnaire is not unidimensional, alpha underestimates reliability (John & Benet-Martinez, 2000). Separate unidimensional subscales should therefore be scored and alpha indexed separately (John & Benet-Martinez, 2000).

Some basic assumptions are made when estimating alpha. These include random sampling of participants and items, and that items are independent in the sense that responding to one item does not affect an individual's performance on subsequent items, as a result of e.g. fatigue or confusion (Cronbach & Shavelson, 2004). One can rarely be sure that these assumptions are not violated, so it is important to be careful when interpreting alpha (Cronbach & Shavelson, 2004). It is also important to remember that reliability estimates such as Cronbach's alpha are a property of the scores of a measure, not the measure itself, and these estimates are therefore said to be sample dependent (Cortina, 1993). The current study examines a community-based sample of adolescents in their early teens and their mothers, and seeks to establish the reliability of the SMFQ in this group.

Content validity. The content validity of a questionnaire describes to what degree it covers the full range of material of interest, and whether the questionnaire adequately reflects the disorder of interest (Angold et al., 1995). As depressive symptomatology is wide-ranging, depression questionnaires can include many different items covering areas such as indecision, suicide, sleep patterns and loneliness. The SMFQ was designed to capture core symptomatology included in the Diagnostic and Statistical Manual of Mental Disorders, 3rd ed. criteria for major depressive disorders (DSM-III; APA, 1980; Angold et al., 1995; Messer et al., 1995). The DSM-III and the DSM-IV (APA, 2000) symptom criteria for major depression are identical (Sund et al., 2001), and very similar to the ICD-10 (WHO, 1992). Through examining whether the SMFQ adequately reflects core depressive symptomatology in both the DSM-IV and the ICD-10, content validity can be assessed.

Symptoms of Depression in Adolescents

Depression in adolescents has similar clinical features and presentation of symptoms as depression in adults, and adolescent depression can predict a variety of mental and physical health problems later in life (Friedman & Anderson, 2014; Thapar, Collishaw, Pine, & Thapar, 2012). Symptoms of depression can vary in severity and composition, and core symptomatology in the DSM-IV and the ICD-10 includes depressed or irritable mood, loss of interest or pleasure, change in psychomotor activity, decreased energy, reduced concentration, sleep and appetite disturbances, feelings of guilt or worthlessness, reduced self-esteem and thoughts of self-harm or suicide (APA, 2000; WHO, 1992; appendix A).

In a longitudinal study, over 50% of cases with first-onset depression had prior depressive symptoms (Horwath, Johnson, Klerman, & Weissman, 1992). Symptoms of depression in adolescence strongly predict depression in adults, also in adolescents with sub-clinical levels of depressive symptomatology, and many adolescents have depressive symptoms under the threshold for diagnosis. (Pine et al., 1999). It has been suggested that prospective epidemiologic studies offer the best design for identifying early predictors of depression, and that early identification of predictors can aid interventions that seek to change the course of a disorder before it becomes chronic (Pine et al., 1999). Depressive symptoms are often unrecognized and untreated therefore identification of individuals with depressive symptoms can represent an important step in the prevention of depression (Horwath et al., 1992). However, in order to provide preventive care, one needs valid and reliable tools for identifying at-risk or affected children and adolescents. Establishing whether the SMFQ is a valid and reliable tool

in a sample of 12-13 year old adolescents and their parents in Norway, is the primary aim of the current study.

The composition and severity of depressive symptoms is typically reported differently among adolescent girls and boys, where girls report more severe depressive symptoms more frequently than boys do (Lundervold et al., 2013). On the other hand, symptoms of depression seem to have more negative consequences for boys than girls. This was illustrated in a study of 13-17 year olds, where self-reported depressive symptoms had a stronger negative impact on the functioning of boys (Derdikman-Eiron et al., 2012). The prevalence rates of depression in adults are approximately twice as high in women as it is in men (Hankin et al., 1998). The trend of higher rates of depression in females starts to emerge in early adolescence (12-15 years), and continues to rise throughout late adolescence (15-18 years) (Hankin et al., 1998). This might suggest that adolescence represents a critical time with the greatest risk for onset of depression, particularly among females (Hankin et al., 1998). The current study will seek to establish whether higher levels of depressive symptomatology in girls is the case also in this sample of 12-13 adolescents.

Reports of depressive symptoms can be gathered from both adolescents and their parents. Parent-child agreement in multi-informant studies is generally low, often with correlations in the .20 - .30 range (Achenbach, McConaughy, & Howell, 1987; De Los Reyes & Kazdin, 2005). It has been suggested that this lack of agreement might be due to the fact that parent reports and child self-reports measure different information, and not that one informant's judgement is invalid or wrong (Karver, 2006). Which informant (parent or child) is being used to provide measures of depressive symptomatology and cognitive processes, can influence whether a diagnosis of depression is given (Kazdin, 1989). The age of children and adolescents can influence agreement, where previous research has indicated that the older children and adolescents are the more parents and their children agree (Renouf & Kovacs, 1994). Prior investigations of the SMFQ from the TOPP study data have indicated a moderate parent-child agreement, with estimated correlations of .49 (Karevold, 2008). The current study seeks to deepen the understanding of the agreement, by separating and comparing the agreement by sex, i.e. comparing mothers of girls with girls, and mothers of boys with boys.

Socioeconomic status (SES) and mental health problems are often inversely related (Reiss, 2013). In a recent review of 52 studies focusing on the relationship between socioeconomic status and mental health problems in children and adolescents, there was an inverse relationship between socioeconomic status and mental health problems in children and adolescents (Reiss, 2013). The relationship between indicators of SES and mental health

problems was found in all age groups, but there were no consistent gender patterns. The strongest predictors of mental health problems in children and adolescents was low household income and low parental education (Reiss, 2013). The current study will examine the relationship between SES indicators and symptoms of depression in a sample of adolescents in their early teens in Norway.

Research Questions

To account for the aforementioned limitations in the research field, the following research questions have been posed and will be examined in the current study.

Psychometric Examinations of the SMFQ

I. Does the SMFQ have structural validity demonstrated through unidimensionality?
II. Does the SMFQ have adequate reliability?
III. Does the SMFQ have content validity, i.e. does it cover core depressive symptomatology?

Secondary Examinations

IV. How much do parents and children agree on reports of depressive symptomatology?
V. What is the threshold of depressive symptomatology indicative of depression?
VI. What are the ratios of girls to boys that score above this threshold?
VII. What is the relationship between socioeconomic status and depressive symptomatology?

Method

Sample and Procedure

The TOPP study. The TOPP study ("Tracking Opportunities and Problems in Childhood and Adolescence" / "Trivsel og Oppvekst i Barndom og Ungdomstid") is a study that investigates developmental pathways to good mental health, mental distress, happiness and alcohol use in children and adolescents. In parents, areas such as mental health, relationships, divorce, and absenteeism are investigated. The purpose of the TOPP study is to gather knowledge about precursors, paths of development and factors of risk and protection with regards to mental distress and disorders, in order to examine negative developmental trajectories and to promote good mental health (Norwegian Institute of Public Health, 2015;

NIPH). An understanding of developmental processes is crucial in order to be able to provide optimal conditions during childhood and adolescence, and knowledge about developmental processes is important when planning preventative care (NIPH, 2015).

The Norwegian Institute of Public Health is currently responsible for the processing of data from the TOPP study. The data is considered unique and valuable, and few studies nationally and internationally, have gathered similarly extensive amount of information from families (NIPH, 2015). In addition to demographical and socio-economical questions, parents were asked questions about issues as wide ranging as social relations with neighbours, health, coping style, and quality of partner relationship. Both parents and children answered questions about the child's social skills, anxiety, depression, temperament, somatic health, stressors and life events and other psychosocial factors of risk and protection in and outside the individual and the family (NIPH, 2015).

The TOPP study gathered information at eight different points in time, from when the children were about 18 months until they were about 18 years. The data collection started in 1993, when mothers visiting 19 different health clinics in eight separate areas of Eastern Norway were asked to fill out a questionnaire packet (NIPH, 2015). When the TOPP study started, more than 95% of families with children in the areas in question, used the health stations (Gustavson, Røysamb, von Soest, Helland, & Mathiesen, 2012; Mathiesen, Sanson, Stoolmiller, & Karevold, 2008). Of the parents that entered the study, 913 were mothers and 16 were fathers. The fathers were eliminated from the analyses. (Gustavson et al., 2012). More than 95% of the people who entered the study were of Norwegian ethnicity (Mathiesen et al., 2008). The age of the mothers ranged from 19 to 46 years (M = 30, SD = 4.7), and 28% of the families lived in large cities, 55% in densely populated areas and 17% in rural areas (Karevold, Coplan, Stoolmiller, & Mathiesen, 2011).

Concerning response rate, about 87% of mothers who were asked to enter, filled out the questionnaire in the first wave (T1), and the response rate fluctuated between 83% and 45% during the study. In the current study, data from the fifth (T5) wave of questionnaires will be used, i.e. when children were 12-13 years (girls 53.4%; boys 46.6%). The SMFQ is the subscale used to gather information regarding depressive symptomatology in the TOPP study. Also included in the current study are questions regarding indicators of socioeconomic status. Table 1 has an overview of the data collection waves in the TOPP study, the highlighted fields are the data used in the current study.

Table 1
Participation in the TOPP study from 1993 to 2011.

Data collection waves	T1	T2	T3	T4	T5	T6	T7	T8
Year	1993	1994	1996	2000	2004	2006	2008	2011
Mean age	1.5	2.5	4.5	8.5	12.5	14.5	16.5	19
N mothers	913	777	727	505	587	474	421	520
% mothers of T1	87%	83%	80%	55%	65%	51%	45%	56.9%
N adolescents	-	-	-	-	566	458	375	442
% adolescents of T1	-	-	-	-	58%	49%	40%	47.9%
N fathers	-	-	-	-	-	369	309	371
% fathers of T1	-	-	-	-	-	39%	33%	40.6%

Attrition. Prior attrition analyses indicate that the sample is still representative of those initially invited into the study, and indicate that the TOPP data overall does not suffer from attritional bias. (NIPH, 2015; Karevold et al. 2011). At the start of the TOPP study, 929 families participated, and the percentages of mothers who still participated at T5 was 65%. The percentage of adolescents who participated at T5 was 58%. This gives percentages of attrition of 35% for mothers and 42% for adolescents, over the course of an 11 year period.

Attrition analyses from T1 to T5 suggest that remaining families do not significantly differ in areas such as maternal distress, family adversities, social support, and child temperament (Karevold, Røysamb, Ystrom & Mathiesen, 2009). The only variables predicting dropout at T5 were years of education, i.e. the more education mothers had, the less likely they were to drop out (Karevold et al., 2009). There were no significant differences between mothers who chose to not participate after T1 and those still in the study at T7, except on level of education in mothers, i.e. those who continued to participate had higher levels of education (Gustavson et al., 2012; Kjeldsen, 2013).

Instruments

SMFQ. The Short Mood and Feelings Questionnaire is an instrument designed for the quick evaluation of depressive symptomatology in children and adolescents (ages 8-18), for use in research and epidemiology, and is meant to capture symptoms included in the DSM-III & DSM-IV criteria for major depressive disorders (Angold et al., 1995; Messer et al., 1995). The DSM-III and the DSM-IV symptom criteria for major depression are identical (Sund et al., 2001). The SMFQ is the short version of the 'Mood and Feelings Questionnaire' (MFQ) (Angold et al., 1987) and there are questionnaires for child self-report and parental report for

both versions. The SMFQ has 13 items, the selection of which was based on a larger pool of 30 items, and was meant to reflect the, then, current clinical and taxonomic thinking about childhood depression. The items that had performed well in a variety of analyses were selected for inclusion in the SMFQ (Angold et al. 1995).

The SMFQ is a widely used tool for studying depression in adolescence, and is an appropriate instrument for use in prospective studies from childhood through late adolescence (Turner et al., 2014). Empirical evidence also suggest that the SMFQ is effective in discriminating between clinically referred psychiatric subjects from unselected paediatric controls, and that it is able to distinguish between depressed and non-depressed subjects in the general population (Angold et al., 1995; Costello & Angold, 1988). It correlates well with the Children's Depression Inventory (CDI; r = .67) and generally high reliability has been documented (Angold et al., 1995; Angold, Erkanli, Silberg, Eaves, & Costello, 2002; Sharp et al., 2006).

The questionnaire has a common response format, where each item is rated on a 3-point Likert scale with values of 0, 1 and 2 assigned to response statements of 'Not True', 'Sometimes' and 'True', respectively, given a timeframe over the last two weeks. Because the sum of the raw scores are used, one can get a sum-score of 0-26 on the SMFQ. A high score translates into a high burden of depressive symptomatology. The long version, the MFQ, has been translated to Norwegian and then reverse translated, and the original author has accepted the final version (Sund et al., 2001). In the TOPP study, 12 of 13 items that originally were in the SMFQ were presented. The item "I found it hard to properly think or concentrate" was excluded, because another scale in the questionnaire packet in the TOPP study already included a similar item. The versions using 12 and 13 items are highly correlated ($r = .99$, $p < .001$) (Nilsen, Karevold, Røysamb, Gustavson, & Mathiesen, 2012).

Both mother's reports on children and child self-report will be used in the analyses, and only the mothers and children who completed the SMFQ at T5 concurrently will be included in the study. This is done in order to prevent issues with missing data. The number of individuals in each group is: Mothers N = 505 and children N = 505 (girls N = 274, 54%; boys N = 231, 46%). Appendix B has an overview of items included in the English and Norwegian versions of the original SMFQ child self-report; appendix C has the Norwegian child self-report form that was given in the TOPP study (12 items); appendix D has the Norwegian mother report on children from the TOPP study.

Indicators of socioeconomic status. The following questions assessing socioeconomic status are included in the current study: Mother's report on own and partner's education, rated on a 5-point scale, ranging from elementary school to four years of university. Mother's report on own and partner's employment status, rated on a 4-point scale, ranging from no gainful employment to full-time work. Mother's report on coping with family economy, rated on a 5-point scale, ranging from very poor to very well. Mother's report on family income, rated on an 8-point scale, ranging from no income to over 600.000 NOK per year. The current study also includes a question regarding relationship status of mothers, i.e. if mothers have a partner or not.

Plan of Analysis

Internal structure. Empirical evidence suggests that depression can be construed as a unidimensional phenomenon (Aggen et al., 2005). The SMFQ was constructed as a unidimensional questionnaire (Angold et al., 1995; Messer et al., 1995) and previous validation studies have analysed the SMFQ as a unidimensional measure (e.g. Angold et al., 2002; Costello et al., 1991; Lundervold et al., 2013 Messer et al., 1995; Sharp et al., 2006; Turner et al., 2014). The internal structure of the SMFQ was therefore examined through analysing the factor structure, and confirmatory factor analyses (CFA) were performed on reports from mothers, girls and boys, both separately and combined.

The SMFQ has a Likert-scale response form, meaning that it produces ordinal data. Ordinal data typically violate the assumptions of the Pearson correlation matrix (Timmerman, & Lorenzo-Seva, 2011). Pearson correlations often underestimate the strength of relationships between ordinal items (Olsson, 1979). Therefore, polychoric correlations were used, which is the recommended method for factor analysis performed on data from ordinal variables (Baglin, 2014). Polychoric correlations are advised when the univariate distributions of ordinal items are asymmetric or have an excess of kurtosis (Muthén & Kaplan, 1985).

There are many different methods of extraction that can be used in CFA, among these are maximum likelihood (ML), weighted least squares (WLS), robust WLS, diagonally WLS (DWLS), unweighted least squares (ULS) and principal axis factoring (PAF) – as well as principal components analysis (PCA). ML assumes that the observed variables are continuous and normally distributed (Bollen, 1989), making it unsuitable in this study. In a Monte Carlo study, comparing WLS and robust WLS in CFA of ordinal variables, the authors conclude that WLS in most cases led to substantial estimation difficulties (Flora & Curran, 2004). In a Monte Carlo study of ULS versus DWLS using ordinal variables, the authors conclude that ULS is

better, as it provides more accurate and less variable parameter estimates, and more precise standard errors (Forero, Maydeu-Olivares & Gallardo-Pujol, 2009). In a Monte Carlo study of the performance of the polychoric correlation coefficient in CFA with ordinal data, ULS avoided many of the major failings of other fitting functions (Rigdon & Ferguson, 1991). There are advantages and drawbacks with all types of extraction, so the extraction method should be based on both the response format and the distribution of the data (Baglin, 2014). After deliberations and comparisons, unweighted least squares was chosen as the extraction method.

Factor analysis is generally considered a subjective exercise (Fabrigar, Wegener, MacCallum & Strahan, 1999), and there are various strategies to estimate how many factors to retain after a factor analysis. A popular method of deciding how many factors to retain, is the parallel analysis (PA; Baglin, 2014, Field, 2013). A PA generates random data based on the same number of items and the number of respondents as in the data material (Horn, 1965). PA minimum rank factor analysis (PA-MRFA) is an alternate implementation of PA. PA-MRFA factor retention criteria are based on the randomly generated explained common variance, and not on eigenvalues as proposed by Horn (1965) (Timmerman & Lorenzo-Seva, 2011). PA-MRFA has performed better than PA based on principal components analysis and PA based on principal axis factoring (Timmerman & Lorenzo-Seva; 2011).

The factor analyses were performed using the FACTOR program v9.3.1 (Lorenzo-Seva, U., & Ferrando, P.J., 2006), set to analyse the polychoric correlations, setting the number of factors to 1, using the ULS extraction method and basing factor retention on PA-MRFA. The program produces three measures of sampling adequacy: Firstly, the Kaiser-Meyer-Olkin (KMO) statistic, which is a summary of how small the partial correlations are relative to the original correlations, on which >.8 is considered good (Cerny & Kaiser, 1977; Field, 2013). Secondly, Bartlett's test of sphericity, which tests whether the correlations in the data set are suitable for factor analysis. This test needs to be significant ($p < .01$) (Field, 2013). Thirdly, it produces the determinant of the correlation matrix, which must be checked to see if it is larger than .00001, in order to be sure that multicollinearity is not a problem (Pallant, 2013; Field, 2013).

Factor loadings, communalities and the prevalence of the two most severe response categories (true & sometimes true) will be presented for each factor analysis. Also included in the results are each factor's percentage of explained common variance, as well as results from the parallel analysis (PA). Results from the PA will be presented as the 95th percentile of each factor's explained common variance, based on randomly generated data. Three model-fit indices will be presented: Bentler's comparative fit index (CFI; Bentler, 1990), the Tucker-

Lewis Index (TLI; Tucker & Lewis, 1973) and the root-mean error of approximation (RMSEA; Steiger & Lind, 1980). Acceptable fit cut-off values will follow the recommendations made by Yu (2002): CFI > 0.96, TLI > 0.95 and RMSEA < 0.05.

Factorial invariance. Factorial invariance across boys and girls, and across children and mothers, was examined by fitting structural equation models with factor loadings constrained as equal, using Mplus (Muthen & Muthen, 2012). Models with equal loadings should show acceptable fit by common criteria, i.e. CFI > 0.96, TLI > 0.95, and RMSEA < 0.05 (Yu, 2002).

Reliability. Cronbach's alpha assumes unidimensionality, so in order for alpha to be an unbiased estimate of population reliability, the SMFQ needs to exhibit unidimensionality (Cortina, 1993). After performing the confirmatory factor analyses to see if there were any departures from unidimensionality, the reliability estimates, given as Cronbach's alpha and mean inter-item correlations were computed separately for mothers, children, boys, girls, mothers of girls, mothers of boys and all combined, using SPSS (version 22).

Content validity. The SMFQ was designed to capture core symptomatology included in the DSM-III-R & DSM-IV criteria for major depressive disorders (Angold et al., 1995; Messer et al., 1995). Each item in the SMFQ should cover at least one area of diagnostic criteria, and conversely, each diagnostic criterion should be covered by at least one item (Carmine & Zeller, 1979; Streiner & Norman, 1989). Content validity was therefore examined through tabulating and comparing items in the questionnaire with the diagnostic criteria for depression in the ICD-10 and in the DSM-IV, in order to surmise whether the SMFQ accurately reflects depressive symptomatology.

Parent-child agreement. The parent-child agreement was estimated through correlating children's sum-scores on the SMFQ with mothers' SMFQ sum-scores, using SPSS (version 22). This was done both combined and separated by sex. When two measures are correlated, measurement error makes the correlation coefficient lower than it would have been, had the measures been precise, according to classical test theory (Muchinsky, 1996). The reliability of measures is the proportion of observed variance not caused by measurement error. Muchinsky (1996) states that measurement error can be removed from a correlation coefficient by correcting for attenuation, and is an estimate of how high the correlation would be if two

variables were perfectly reliable. Muchinsky goes on to state that correcting for attenuation does not account for sampling error and a correlation corrected for attenuation cannot be tested for significance. A correlation corrected for attenuation does not increase the predictive value of a measure, and normal correlations and disattuenuated correlations cannot be compared directly. There seems to be debate whether to correct correlations for attenuation or not, but it is recommended to provide both the corrected and uncorrected coefficients (See Muchinsky, 1996 for a review).

Threshold of depression, ratios & comparisons. A threshold of depressive symptomatology indicative of depression can be defined as a cut-off score. Depending on the instrument used, a cut-off score can indicate diagnosis of disorder, operationalisation of disorder, further questioning, or pass or fail. In the case of the SMFQ, which was designed for the quick evaluation of depressive symptoms (Angold et al. 1995), a cut-off above a certain point is meant to signify a burden of depressive symptomatology indicative of depression. Provided that findings support a unidimensional interpretation of the SMFQ, the total SMFQ score can be used as a continuous variable representing severity of depressive symptomatology.

The SMFQ is not a diagnostic instrument per se; therefore a score above a set cut-off, which is different in each sample, could instead be used as a starting point for further investigation. In an ideal setting, all individuals who are administered the SMFQ are also administered a diagnostic interview or a diagnostic instrument, through which one can surmise whether the individual in question can be diagnosed with depression. In the TOPP study there were no diagnostic interviews with which to corroborate a chosen cut-off point. Therefore, one cannot say that a score above a certain point implies a clinical diagnosis of depression.

Different studies have found different cut-off scores: A cut-off score of 8 or more on the SMFQ had the best sensitivity and specificity compared to a depression diagnosis in a general population sample, in a study of psychiatric and unselected paediatric controls aged 8-16, (Angold et al., 1995). In a study of two large, longitudinal samples of twins and singletons aged 8-17 the cut-off point was set at 11, which selected the top 6% of SMFQ scores (Angold et al., 2002). A large population-based study of adolescents aged 16–18 years in Western Norway, used the 90th percentile score as an operationalisation of depression, because there is no valid cut-off score for adolescents in Norway (Sivertsen, Harvey, Lundervold & Hysing, 2014). One of the authors of the SMFQ advises that there is no single cut point that is best for use in all circumstances, but that it is instead up to users to decide what will be most useful in their particular circumstances (Angold, 2008).

In light of previous research and the fact that there does not exist a valid cut-off score for Norwegian adolescents, the cut-off in the current study was set at the 90th percentile, and was used as an operationalisation of depression. In other words, 90% of individuals will have sum-scores below the cut-off and 10% will have sum-scores above. The cut-off points were calculated for girls and boys, as well as for the mother reports, using SPSS (version 22). Mean scores were compared across the various applicable groups to see if there were any statistically significant differences in mean scores, using SPSS (version 22). The ratios of girls to boys that score above the chosen cut-off point were calculated using the aforementioned operationalisation of depression.

What is the relationship between socioeconomic status and depressive symptomatology?
The hypothesized inverse relationship between indicators of socioeconomic status and sum-scores on the SMFQ was examined through multiple one-way ANOVA's, using SPSS (version 22). The one-way ANOVA's were performed for both boys' and girls' SMFQ sum-scores, and were compared in a stepwise manner with the seven different SES indicators, using a Games-Howell post-hoc analysis and testing for homogeneity of variance. A post-hoc analysis was done because no specific hypotheses were made beforehand, variances were assumed unequal and group sizes were different (Field, 2013). By utilising the Games-Howell post-hoc analysis method, Type I errors are less likely to occur, and the multiple comparison family-wise error rate one can be controlled (Field, 2013). Any effects found are described and correlated to examine the direction and effect the SES indicators were thought to have on SMFQ scores.

Results

Participant Characteristics.

Only cases in which both the children's report and mothers' report of the SMFQ were completed concurrently were included in the study. This gives mothers N = 505; children N = 505; girls N = 274 (54.3%); and boys N = 231 (45.7%). The children's mean age was 12.5 years.

There were seven indicators of socioeconomic status. Regarding highest level of mother's education, the majority (52.8%) were well educated, with one to four years or more of college or university. 42.1% had one to three years of high school, and 5.1% had only elementary school. The education levels were similar for partners, with 58.5% having attended college or university for one to four years or more. About a third (35.5%) of partners had one

to three years of high school, and 6% had only elementary school. Regarding mother's employment status, most mothers (55.2%) had full-time work, and about a third (31.2%) had part-time work, while 13.6% had no gainful employment. The rates of employment for partners were higher, where 89.2% had full-time work, 2.3% had part-time work and 2.3% had no gainful employment (5.5% had no partner). About three quarters of mothers (76.6%) were coping well or extremely well with the family economy, about a fifth of mothers (20.9%) said they were coping OK, while 2.6% of mothers said the economy was either poor or very poor. Regarding total family income in NOK, 72.2% had between 400 thousand and 600 thousand or more, 21.6% had between 200 thousand and 400 thousand, while 6.1% had under 200 thousand in total family income. Most mothers (94.5%) had a partner, while 5.5% of mothers were single.

Psychometric Examinations

Factor analysis. Results from the factor analyses are presented in tables 2-6, and show the results for all combined, children, mothers, girls and boys, respectively. Items, factor loadings, communalities, prevalence of true and sometimes true, the factors, percentages of explained common variance and parallel analysis percentages of explained common variance (given as the 95th percentile of variance) are presented. The determinant of the correlation matrix, whether Bartlett's test was significant, the Kaiser-Meyer-Olkin statistic, the CFI, the TLI and the RMSEA are also presented.

All combined.

Table 2

Results from factor analysis of the SMFQ – Mothers & children combined

Items	*Factor loadings*	*Commu-nalities*	*Prev. of true & sometimes true*	*Factor*	*% of variance*	*PA % of variance*
Miserable or unhappy	.735	.541	40.1%	1	61.9	19.4
Felt tired	.480	.231	39.6%	2	8.5	17.1
Very restless	.336	.113	32.0%	3	7.4	15.1
Didn't enjoy anything	.688	.474	11.8%	4	5.0	13.3
Felt no good	.840	.706	17.5%	5	4.1	11.7
Cried a lot	.715	.511	13.0%	6	3.3	10.2
Hated him/herself	.854	.728	11.0%	7	3.1	8.8
Never be as good	.746	.556	18.5%	8	2.7	7.4
Felt lonely	.702	.493	20.3%	9	2.2	6.4
Nobody loved him/her	.824	.679	10.1%	10	1.7	4.8
Bad person	.847	.717	10.7%	11	0.2	3.3
Did everything wrong	.804	.646	21.3%	12	0.0	0.0

N = 1010; Determinant = .027; Bartlett's sig ($p < .001$); KMO = .909

CFI = .98; TLI = .98; RMSEA = .039

Children.

Table 3

Results from factor analysis of the SMFQ - Children

Items	*Factor loadings*	*Commu-nalities*	*Prev. of true & sometimes true*	*Factor*	*% of variance*	*PA % of variance*
Miserable or unhappy	.721	.520	41.4%	1	60.6	19.0
Felt tired	.463	.215	48.3%	2	8.8	16.5
Very restless	.279	.078	45.9%	3	7.5	14.6
Didn't enjoy anything	.654	.428	13.7%	4	6.6	12.9
Felt no good	.816	.666	18.4%	5	5.5	11.5
Cried a lot	.701	.491	16.0%	6	3.5	10.1
Hated myself	.845	.715	13.5%	7	2.5	8.8
Never be as good	.766	.587	20.6%	8	2.1	7.5
Felt lonely	.753	.568	21.6%	9	1.4	6.1
Nobody loved me	.793	.629	11.9%	10	1.0	4.7
Bad person	.825	.680	12.9%	11	0.5	3.1
Did everything wrong	.848	.718	19.2%	12	0.0	0.0

N = 505; Determinant = .023; Bartlett's sig ($p < .001$); KMO =.899

CFI = .98; TLI = .98; RMSEA = .043

Mothers.

Table 4

Results from factor analysis of the SMFQ - Mothers

Items	*Factor loadings*	*Commu-nalities*	*Prev. of true & sometimes true*	*Factor*	*% of variance*	*PA % of variance*
Miserable or unhappy	.748	.559	38.8%	1	61.0	23.0
Felt tired	.478	.228	30.9%	2	8.7	16.4
Very restless	.398	.158	18.0%	3	7.6	14.0
Didn't enjoy anything	.737	.544	9.9%	4	6.2	12.5
Felt no good	.878	.771	16.6%	5	5.8	11.0
Cried a lot	.725	.525	9.9%	6	3.3	9.6
Hated him/herself	.865	.749	8.5%	7	2.4	8.4
Never be as good	.704	.495	16.4%	8	2.0	7.1
Felt lonely	.628	.395	19.0%	9	1.6	6.0
Nobody loved him/her	.869	.756	8.3%	10	1.2	4.5
Bad person	.877	.769	8.5%	11	0.2	3.0
Did everything wrong	.781	.610	23.4%	12	0.0	0.0

N = 505; Determinant = .032; Bartlett's sig ($p < .001$); KMO = .894

CFI = .99; TLI = .98; RMSEA = .042

Girls.

Table 5

Results from factor analysis of the SMFQ - Girls

Items	*Factor loadings*	*Commu-nalities*	*Prev. of true & sometimes true*	*Factor*	*% of variance*	*PA % of variance*
Miserable or unhappy	.749	.561	47.9%	1	58.9	20.3
Felt tired	.442	.195	53.1%	2	9.4	16.9
Very restless	.311	.097	43.8%	3	7.2	14.8
Didn't enjoy anything	.743	.553	13.0%	4	6.5	13.0
Felt no good	.812	.659	20.2%	5	4.8	11.4
Cried a lot	.764	.583	21.6%	6	4.1	10.0
Hated myself	.825	.681	16.8%	7	3.2	8.6
Never be as good	.733	.538	20.5%	8	2.9	7.4
Felt lonely	.724	.524	23.3%	9	1.3	6.0
Nobody loved me	.783	.613	12.3%	10	1.2	4.6
Bad person	.787	.620	14.0%	11	0.4	3.0
Did everything wrong	.865	.748	22.3%	12	0.0	0.0

N = 274; Determinant = .016; Bartlett's sig ($p < .001$); KMO =.883

CFI = .98; TLI = .98; RMSEA = .040

Boys.

Table 6

Results from factor analysis of the SMFQ - Boys T5

Items	*Factor loadings*	*Commu-nalities*	*Prev. of true & sometimes true*	*Factor*	*% of variance*	*PA % of variance*
Miserable or unhappy	.653	.427	33.3%	1	57.3	23.8
Felt tired	.466	.217	44.4%	2	9.1	16.8
Very restless	.258	.066	48.6%	3	8.9	14.2
Didn't enjoy anything	.579	.336	13.6%	4	7.5	12.4
Felt no good	.810	.655	16.0%	5	6.1	11.0
Cried a lot	.605	.366	8.6%	6	3.8	9.4
Hated myself	.868	.753	9.1%	7	2.9	8.2
Never be as good	.821	.675	19.8%	8	1.5	7.1
Felt lonely	.802	.644	18.9%	9	1.4	5.8
Nobody loved me	.827	.684	10.7%	10	1.2	4.5
Bad person	.880	.774	11.1%	11	0.1	3.2
Did everything wrong	.799	.638	14.8%	12	0.0	0.0

N = 231; Determinant = .016; Bartlett's sig ($p < .001$); KMO = .865

CFI = .96; TLI = .95; RMSEA = .048

Most items had both skewness and kurtosis higher in absolute value than 1, indicating that choosing to use polychoric correlations was appropriate (Muthén & Kaplan, 1985). Regarding sampling adequacy, we can see that all KMO statistics are over the required .8, with a range of .865 to.909. All Bartlett's tests are significant, and all determinants are above the required .00001. Combined, this indicates that the data was suitable for factor analysis.

All model-fit indices fell within acceptable ranges (CFI over 0.96, TLI over 0.95 and RMSEA under 0.05; (Yu, 2002)), with CFI's between .96 and .99, TLI's between .95 and .98, and RMSEA's between .039 and .048. CFI is based on a comparison between a baseline model and a hypothesised model, and as this index considers sample size, it generally avoids underestimation (Bentler, 1990; Byrne, 2001). TLI also takes sample size into account (Byrne, 2001; Tucker & Lewis, 1973). Values above .95 on the CFI and the TLI indicate good fit (Hu & Bentler, 1999). RMSEA attempts to correct the chi-square statistic's tendency to reject models with a large N or many observed variables, and values lower than .05 indicate good fit (Byrne, 1990; Steiger & Lind, 1980).

Factorial invariance. Results from examinations of factorial invariance are shown in table 7. Chi-square values and degrees of freedom, as well as CFI, TLI and RMSEA are reported.

Table 7
Factorial invariance

	Mothers & Children		*Boys & Girls*	
	Loadings free	*Loadings equal*	*Loadings free*	*Loadings equal*
Chi-square	320.74	310.86	170.69	193.31
df	251	263	119	131
CFI	.98	.99	.97	.97
TLI	.98	.99	.97	.97
RMSEA	.02	.02	.04	.04

Reliability. As shown in table 8, the reliability estimates range from α =.80 to α = .85. The mean inter-item correlations range from r = .28 to r = .34. The separated scales of girls and mothers of girls had the highest reliability and mean inter-item correlation, both with α =.85 and r = .34. The mothers of boys' separated scale has the lowest reliability and mean inter-item correlations, with α =.80 and r = .28.

Table 8
Cronbach's alpha and mean inter-item correlations (Miic)

Group	*N*	*Cronbach's alpha*	*Miic*
All combined	1010	.84	.33
Children	505	.84	.33
Mothers	505	.83	.31
Girls	274	.85	.34
Boys	231	.83	.31
Mothers of girls	274	.85	.34
Mothers of boys	231	.80	.28

Content validity. The items in the SMFQ appear to cover the majority of the core symptomatology in the ICD-10 (WHO, 1992) and DSM-IV (APA, 2000) (see appendices A & B). There are ten depressive symptoms in the ICD-10 and nine in the DSM-IV. Seven of these symptoms are covered by at least one item in the SMFQ. The symptom expressing low self-worth appears to be illustrated by the most items in the SMFQ, with a representation of five.

Three depression symptoms are not represented by any SMFQ items, namely disturbed sleep, change in appetite and thoughts of suicide or self-harm. In the SMFQ that was administered in the TOPP study, the item "I found it hard to properly think or concentrate" is not present, but is covered by both ICD-10 and DSM-IV. The two most important symptoms of depression in the DSM-IV, depressed mood and diminished interest or pleasure, and the three most important symptoms in the ICD-10, depressed mood, loss of interest and enjoyment, and increased fatiguability, are covered by SMFQ items.

Secondary examinations

Parent-child agreement. Having found support for the hypothesis that the SMFQ measures a unidimensional construct, the SMFQ sum-score can be used as an indicator of severity of depressive symptoms. SPSS (version 22) was used to analyse the correlation between the children's sum scores on the SMFQ and sum-scores of mother's SMFQ reports, both combined and separated by sex. The correlations were also disattenuated, through dividing the observed correlation by the square root of the product of the two reliabilities (Spearman, 1910). The results of these analyses are presented in table 9.

Table 9
SMFQ Sum-score correlations

Comparison	*N*	*Correlation (r)*	*Sig. (p)*	*Disattenuated correlations (r)*
Children's reports Mothers' reports	505	.44	< .0001	.53
Girls' self-report Mothers' reports on girls	274	.51	< .0001	.60
Boys' self-report Mothers' reports on boys	231	.35	< .0001	.43

Cut-off score. After making sum-scores of children's SMFQ self-reports and mothers' reports, descriptive statistics on percentiles were ordered in SPSS (version 22). Those who had a sum-score equal to or larger than the cut-off scores were classified as depressed. Table 10 has an overview over cut-off scores and the number of individuals operationalised as depressed.

The table shows numbers of self-reported depressed, numbers of reported as depressed by mothers, both children combined and separated by sex.

Table 10

Overview over score means & cut-off scores for the SMFQ.

Group	*N*	*Score Mean*	*SD*	*Cut-off score*	*N depressed*	*Highest Score*
Girls	274	3.9	4.2	11	27	22
Boys	231	3.0	3.5	8	20	19
Mothers of girls	274	2.0	2.8	7	20	17
Mothers of boys	231	2.0	2.7	6	17	17
Children combined	505	3.5	3.9	10	45	22
Mother reports combined	505	2.0	2.7	6	46	17

As one can see from table 10, the cut-off scores differ depending on group. The cut-off scores for girls was estimated to be 11, whereas for boys it was estimated to be 8. For mothers' reports on girls, the cut-off score was estimated to be 7 for girls and 6 for boys.

The combined children's self-reports had statistically significant higher means (M = 3.5, SD = 3.9) than the combined mothers' reports (M = 2.0, SD = 2.7), t = 9.0, p < .0001, with a small to moderate effect size (d = 0.43; Cohen, 1988).

Girls' self-reports had a statistically significant higher mean score (M = 3.9, SD = 4.2) than boys' self-reports (M = 3.0, SD = 3.5), t = 2.6, p = .0104, with a small effect size (d = 0.23; Cohen, 1988).

There was no statistically significant mean difference in mean score between reports from mothers of girls (M = 2.0, SD = 2.8) and reports from mothers of boys (M = 2.0, SD = 2.7).

Girls' self-reports had statistically significant higher means (M = 3.9, SD = 4.2) than mothers' reports on girls (M = 2.0, SD = 2.8), t = 8.4, p < .0001, with a moderate effect size (d = 0.53; Cohen, 1988).

Boys' self-reports had statistically significant higher means (M = 3.0, SD = 3.5) than mothers' reports on boys (M = 2.0, SD = 2.7), t = 4.1, p < .0001, with a small effect size (d = 0.32; Cohen, 1988).

Figure 1, on the next page, shows the distributions of SMFQ sum-scores in the various groups. The SMFQ sum-score represents the horizontal line and the number of individuals the vertical line, for the various groups.

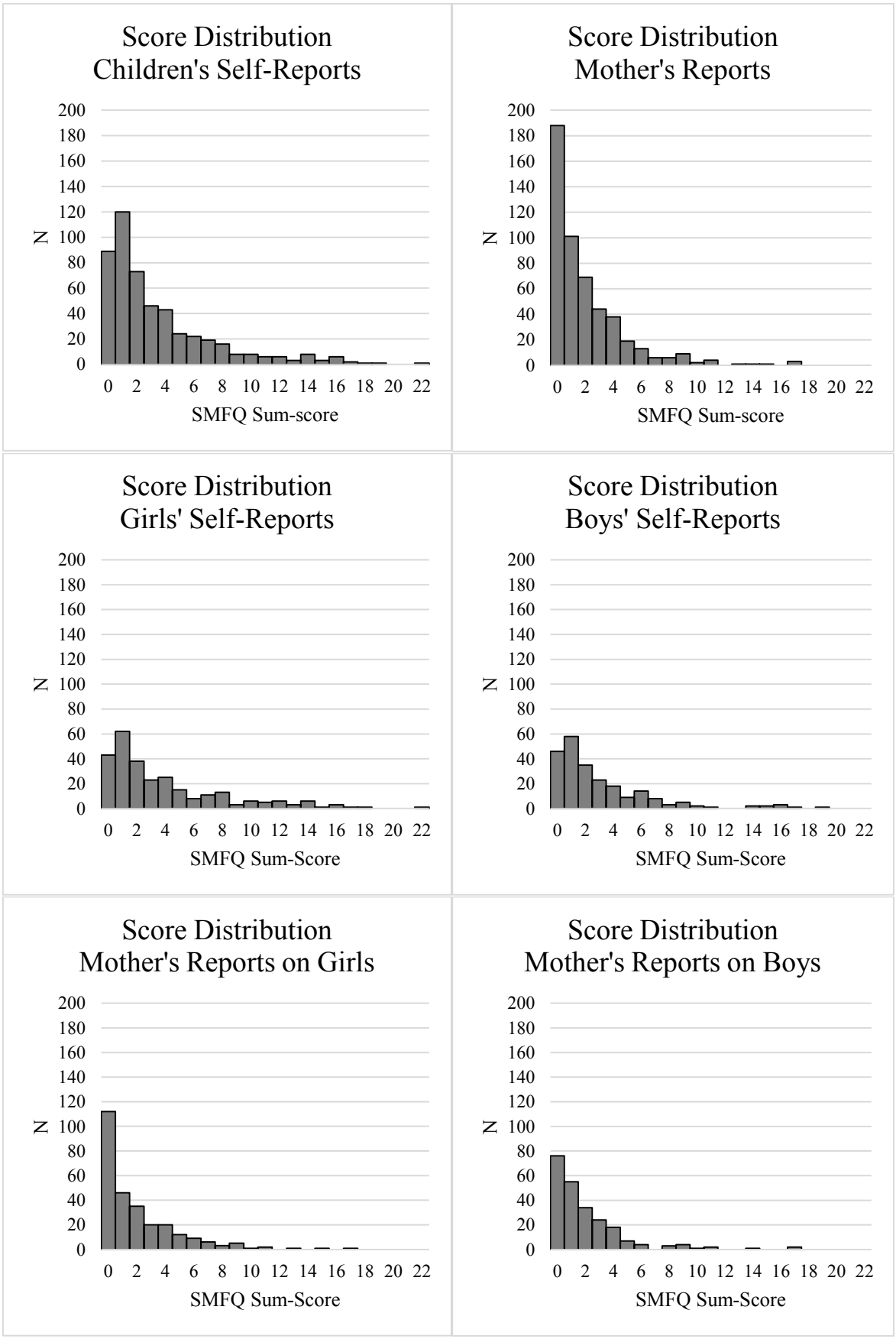

Figure 1 – Score distributions of SMFQ sum-scores in the various groups. SMFQ sum-score represents the horizontal line and the number of individuals the vertical line.

What is the ratio of girls to boys that score above the threshold for depression? The ratio of 12-13 year old girls to boys in this sample, that are classified as depressed is 27 girls to 20 boys or about 4 to 3. In other words, for every 4 depressed girls in this sample of 12-13 year old adolescents, there were 3 depressed boys. For mothers of girls and mothers of boys the ratios were 20 girls to 17 boys. In other words, mothers of girls reported 20 girls to be depressed, whereas mothers of boys reported 17 boys to be depressed.

Socioeconomic status. Multiple one-way ANOVAs were performed in a stepwise manner, as previously described. (The complete results from these analyses are found in appendix E and appendix F). There was significant effect of mothers' relationship status ($F(1, 229) = 5.85$, $p = .016$, $\eta^2 = 0.025$) on boys' SMFQ sum-score. Eta squared of 0.025 is a small effect size (Cohen, 1988), and means that 2.5% of the variance in the boys' SMFQ sum-scores is accounted for by mothers' relationship status. This correlation was positive ($r = .16$, $p = .016$), meaning that sons of mothers who were single, had higher SMFQ sum-scores. There were no other significant main effects nor interaction effects when comparing either of the indicators of SES with boys' SMFQ sum-scores.

For girls there were several effects of SES indicators on SMFQ sum-scores. There was a main effect of family income on girls' SMFQ sum-scores, ($F(6, 262) = 2.46$, $p = .025$, $\eta^2 = 0.053$). Eta squared of 0.053 is a small effect size (Cohen, 1988), and means that 5.3% of the variance in the girls' SMFQ sum-scores is accounted for by family income. This correlation was negative ($r = -.13$; $p = .028$), which means that less income means higher SMFQ scores for girls.

There was a significant main effect of coping with economy on girls' SMFQ sum-scores, ($F(4, 264) = 3.20$, $p = .014$, $\eta^2 = 0.046$). Eta squared of 0.046 is a small effect size (Cohen, 1988), and means that 4.6% of the variance in the girls' SMFQ sum-scores is accounted for by how well the mothers cope with the economy. This correlation was negative ($r = -.16$; $p = .010$), which means that an unsatisfactory family economy means higher SMFQ scores for girls.

In addition there was a significant interaction effect on girls' SMFQ sum-scores, between when partners had less than 50% part-time work and full-time work (part-time mean score = 0, full time mean score = 3.8, mean difference 3.8, $p < .0001$), but there was no main effect in this instance.

Discussion

The aims of the current study were twofold. Firstly, a psychometric examination was undertaken to validate the SMFQ. This was done through examining whether the SMFQ exhibited similar properties as the latent construct depression, namely unidimensionality. Estimations of reliability were then calculated, in order to establish the measurement accuracy of the SMFQ. The content validity of the SMFQ was assessed through comparing items in the SMFQ with depressive symptoms in the DSM-IV and ICD-10. The second aim of the current study was to see if the use of SMFQ sum-scores as a continuous measure of severity of depressive symptoms could be supported. Through this continuous measure gender differences in depressive symptomology, parent-child agreement and ratios of operationalised depression were examined. Lastly, indicators of socioeconomic status (SES) were compared with sum-scores on the SMFQ, in order to see if there was an inverse relationship between SES and depressive symptomology. The following section will discuss the research questions one by one.

Does the SMFQ have structural validity, demonstrated through unidimensionality? Messick (1989) stated that the primary concern in validation is assessing construct validity through examining to what degree an instrument represents the underlying construct. One way to assess this is to examine the internal structure of the SMFQ. The underlying construct in the context of the current study is depression, and depression has been interpreted as being unidimensional (Aggen et al., 2005). The SMFQ was designed to be unidimensional (Angold et al., 1995; Messer et al., 1995). Prior research supports the notion that the SMFQ measures a unidimensional construct (e.g. Angold et al., 1995; Angold et al., 2002; Costello et al., 1991; Lundervold et al., 2013; Messer et al., 1995; Sharp et al., 2006). The SMFQ is scored by summing all item responses to a raw sum-score (Angold et al., 1995), and this sum-score is only a valid measure of depressive symptomatology if the SMFQ is unidimensional.

The current study used confirmatory factor analyses to examine the patterns of symptom covariation. All confirmatory factor analyses support a unidimensional interpretation of the SMFQ. Through comparing the last two columns in each table from the factor analyses (tables 2-6), we can see that the percentages of explained common variance for the first factor in the data is larger than the percentages of random explained common variance for the first factor from the parallel analysis. The second factor is in all instances smaller in the data used in the current study, than randomly generated data from the parallel analysis. This indicates that an

interpretation of a unidimensional internal structure can be supported. This interpretation is also supported by all model-fit indices falling within acceptable values and generally high loadings to the general factor. This means that the current study replicates findings across cultures and languages (Angold et al., 1995; Angold et al., 2002; Costello et al., 1991; Lundervold et al., 2013; Messer et al., 1995; Sharp et al., 2006).

As shown in tables 2-6, most items had high factor loadings to the general factor, the only exception being the item "I was very restless". This item had a low factor loading in both sexes: .311 in girls and .258 in boys, giving it a communality of .097 in girls and .066 in boys. The item had a somewhat higher factor loading for mothers, .398, with a communality of .158. Considering that this item is one of the most endorsed items in children (45.9% of true & sometimes true), this seems rather conspicuous. In other words, many adolescents say that they are restless, but the item has a combined communality of .078 in children, which can be said to be very low. It can be questioned whether restlessness functions as a symptom of depression in this sample. The same might be said for the item "I felt tired", which, in children, has a factor loading of .463 and a communality of .215, but a combined endorsement of 48.3%. Additional reliability analyses also indicate that these two items are the only to increase the overall reliability of the SMFQ, if deleted. The eventual removal of these items from the SMFQ should, however, not simply be based on a reliability analysis, but also a consideration of content validity and the fact that the performance of these items might be better in older groups.

Factorial invariance. Table 7 shows results from examinations of factorial invariance, and indicate that there is factorial invariance across groups. An examination of factorial invariance is done in order to be able to make valid comparisons (Meredith & Teresi, 2006; Wu, Li & Zumbo, 2007).

Does the SMFQ have adequate realibility? Having found support for the notion that the SMFQ measures a unidimensional construct, reliability estimates in the form of Cronbach's alpha were calculated. Considering that the SMFQ has relatively few items (12), and that all reliability estimates fall between α =.80 and α = .85, the reliability estimates of all groups can be considered good (Cortina, 1993). There are, however, differing opinions when it comes to interpreting alpha. Clark & Watson (1995) noted that there are no clear standards as to what level of alpha is considered acceptable.

The mean inter-item correlations are also reported in the current study. The mean inter-item correlations are an estimate of reliability, and can be interpreted as an estimate of the

homogeneity of items in a questionnaire (Briggs & Cheek, 1986). It has been noted that the optimal level of homogeneity is having mean inter-item correlations between r = .2 and r = .4 (Briggs & Cheek, 1986), and that they should fall between r = .15 and r = .50 (Clark & Watson, 1995). All mean inter-item correlations of the different groups in the current study fall between r = .28 and r = .34. The current study finds that the SMFQ has adequate reliability.

Does the SMFQ have content validity, i.e. does it cover core depressive symptomatology? In the assessment of content validity, seven of the core symptoms in the ICD-10 (WHO, 1992) and DSM-IV (APA, 2000) were found in the SMFQ. In the DSM-IV there are two core symptoms, at least one of which needs to be present for a diagnosis of depression to be given. These two core symptoms are depressed mood and diminished interest. In the ICD-10, there are three core symptoms, of which at least two need to be present for a diagnosis of depression to be given. The three core symptoms in the ICD-10 include the two symptoms in the DSM-IV, as well as increased fatiguability. All three are included in the SMFQ, which can be argued to be a strength of the content validity of the SMFQ.

The fact that the TOPP-study version of the SMFQ with 12 items, the item "I found it hard to properly think or concentrate" is not included, can make one to question the content validity. The two different versions are, however, highly correlated (r = .99; Nilsen et al., 2012). Disturbed sleep, change in appetite and thoughts of self-harm or suicide are symptoms not covered by the SMFQ. These three symptoms are in both the DSM-IV and the ICD-10, but because they are not in the SMFQ, these symptoms and their prevalence cannot be investigated using the SMFQ. This can be said to be a weakness in the context of content validity.

In a study of adolescents aged 16-19 in Western Norway, all items had factor loadings of >.60, except restlessness which had .38 in girls and .48 in boys (Lundervold, 2103). The current study replicates the low factor loadings of restlessness, where girls had .31 and boys .26. Removing this item can be argued to not worsen content validity, as it unlikely that this item is functioning as a symptom of depression in this sample of 12-13 year old adolescents.

Questionnaires though to examine depression can include a multitude of items, or symptoms, covering different areas. The SMFQ was specifically and empirically designed to minimize length and maximize criterion validity, i.e. its purpose was not to cover the full range of symptomatology, but rather to be an efficient screening measure (Angold et al., 1995). As such, the current study can conclude that the SMFQ has adequate content validity. This replicates findings across cultures (Turner et al., 2014).

How much do parents and children agree on reports of depressive symptomatology? The current study examined the parent-child agreement through correlating the raw sum-scores on the SMFQ. The overall parent-child agreement was r = .44 (disattenuated r = .53). In itself this is a moderate correlation, but considering that other studies have found correlations as low as .20-.30 (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005), the correlations in the current study are higher. Prior examinations of the parent-child agreement on the SMFQ in the TOPP-study, have revealed a correlation of r = .49 between mothers and children (Karevold, 2008). This estimation of agreement was, however, found using a different methodology. To further investigate the parent-child agreement groups were separated by sex, which revealed that mothers of girls and mothers of boys differed in their agreement with their children. Mothers of girls and girls had a sum-score correlation of r = .51 (disattenuated r = .60), whereas mothers of boys and boys had a sum-score correlation of r = .35 (disattenuated r = .43). In other words, mothers of girls agree more with their children than mothers of boys do.

Overall, the parent-child agreement seems to be moderately higher than prior studies, which have indicated correlations between .20 - .30 (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005). The separation of agreement by sex also represents a new perspective on the TOPP-study data. Topics of further research includes analysing the parent-child agreement as children get older, where one would expect to see more agreement the older the children and adolescents are (Renouf & Kovacs, 1994). The TOPP-study did not include paternal reports until later time points, but the comparison of fathers of boys and mothers of girls, and vice versa, would also be interesting topics for further research.

What is the threshold of depressive symptomatology indicative of depression? When calculating sum-scores and estimating percentiles, the cut-off scores for girls was 11, which classified 27 girls as depressed. The cut-off scores for boys were 8, which classified 20 boys as depressed. For mothers of girls, and mothers of boys the ratios were 27 to 20. These numbers and comparisons of item endorsement rates, indicate that girls in score higher than boys, that more girls are depressed and that girls use the most severe response categories more frequently. This replicates a study of older adolescent (16-19 years), where girls scored significantly higher than boys on the SMFQ, and used true & sometimes true much more frequently (Lundervold et al., 2013).

In the current study, there were no diagnostic instruments or interviews with which to corroborate a chosen cut-off. However, the use of the 90th percentile can be supported considering that about 15-20% of Norwegian adolescents have considerable symptoms of

depression, and that about 5% have depressive symptomology indicative of clinical depression (Sund, 2004; Wichstrøm, 1999). The operationalisation is, however, essentially arbitrary, and a score above a certain cut-off is not indicative of a clinical diagnosis of depression. A score above the cut-off is, on the other hand, indicative of a high burden of depressive symptomatology. Symptoms of depression in adolescence strongly predict depression in adults, also in adolescents with sub-clinical levels of depressive symptomatology, and many adolescents have depressive symptoms under the threshold for diagnosis. (Pine et al., 1999). This could mean that many of the adolescents in the current study have a risk of clinical depression, due to their burden of depressive symptomology. It has been theorized, in a study of depressive symptomatology in Norwegian adolescents, that most of adolescents who have symptoms of depression have problems related to everyday functioning (Lundervold et al., 2013).

What are the ratios of girls to boys that score above the chosen cut-off point? The ratio of girls to boys classified as depressed was 4 to 3. In other words, for every 4 depressed girls in this sample of 12-13 year old adolescents, there are 3 depressed boys. In adults, the ratio is about 2 to 1, i.e. the prevalence of depression is twice as high in women (Hankin et al., 1998). The ratio of girls to boys with adolescent depression varies from country to country. The female to male ratios from different studies varies between 1.3 to 1 (13-18 years; USA), 1.9 to 1 (14-17 years; Germany), 4.1:1 (16-17 years; Sweden) and 8.9 to 1 (7-16 years; Switzerland) (Costello, Erkanli & Angold, 2006; Oldehinkel, Wittchen & Schuster, 1999; Olsson & von Knorring, 1999; Steinhausen & Winkler, 2003). In a recent study in Norway, the girls to boys ratio was 11.8:1 (mean age 15 years) for current MDD, higher than other comparable European countries, attributable in part to a low rate of MDD among boys, but also to methodological constraints (Sund et al., 2011).

Depression in adolescents is often unrecognized, untreated and undiagnosed (Thapar et al., 2012). This seems to also be the case in Norway, where adolescents with depression rarely get help specifically directed at depressive disorders (Sund et al., 2011). It has been suggested that adolescence represents a critical time when the risk of onset of depression is the greatest, particularly among females (Hankin et al., 1998). It has been questioned whether the more affective response-styles in girls than boys, add to reports of more severe depressive symptoms in girls (Lundervold et al., 2013). In addition, it has been theorized that the presence of irritability, mood reactivity and shifting symptoms that are common in adolescence is part of the explanation behind the lacking recognition, treatment and diagnosing of depression in

adolescents (Thapar et al., 2012). This interpretation might be plausible, considering that feelings of misery and unhappiness generally are a part of normal development (Lundervold et al., 2013).

In the current study girls and boys reported depressive symptomatology differently. A proposed explanation behind the difference in reports between sexes, is that girls typically ruminate more than boys. This rumination can in turn lead to a passive and repeated focus on the origins and effects of depressive symptoms (Lundervold et al., 2013). Rumination has been indicated to increase the susceptibility to develop depressive symptoms, and to increase the duration and severity of existing depressive symptoms (Watkins & Moulds, 2005; Abela & Hankin, 2011). On the other hand, it is suggested that the presence of lower prevalence rates of depression in Europe in general, might reflect lower rates among boys in particular, thereby skewing the ratios towards more female depression (Sund et al., 2011). The trend of higher depression rates in women starts to emerge in early adolescence and continues to rise throughout adolescence (Hankin et al., 1998). It can be argued that the trend of higher rates of female depression is starting to emerge also in this sample of 12-13 year old adolescents. A topic of further research could be to examine data from the other time points in the TOPP-study, to find the rates of depression when the children are older.

What is the relationship between socioeconomic status and depressive symptomatology? In short, the main effects were higher levels of depressive symptoms in boys who had single mothers, and in girls living in families with lower income and an unsatisfactory economy. In addition, there was a significant interaction effect on girls' SMFQ sum-scores, between when partners of mothers had less than 50% part-time work and full-time work. This interaction effect is likely a spurious relationship, but a possible interpretation could be that partners spending more time at home contribute positively to the mental health of the girls in question.

In a recent review the strongest predictors of mental health problems in children and adolescents was low household income and low parental education (Reiss, 2013). In the current study, parental education had no significant effects on depressive symptoms. This could be due to effects of attrition. In attrition analyses, the only variables predicting dropout at T5 were years of education, i.e. the more education mothers had, the less likely they were to drop out (Karevold et al., 2009). Considering that mothers who dropped out overall had lower education than mothers who continued, and that there is an inverse relationship between education and mental health (Reiss, 2013), the children of dropouts might have higher levels of depressive

symptomatology than those still in the study. This would represent a bias, and would therefore make generalisations more difficult.

The findings from the current study are in line with findings from other studies, namely that there is an inverse relationship between socioeconomic status and mental health problems (Reiss, 2013). In the current study, girls in families with lower income and an unsatisfactory economy had more depressive symptoms than girls living in families with higher socioeconomic status. Combined with the fact that boys with single mothers had more depressive symptoms, this opens up interesting topics for further research. Possible investigations include whether the inverse effect persists through later time points, and whether a change in the mothers' relationship status have a relation to depressive symptomology in boys.

Strengths and Limitations

Strengths. One of the strengths of this study is that is a community-based sample of adolescents, while many other studies of the SMFQ have selected samples (Turner et al., 2014). The use of community-based or population-based samples enables one to generalise results to a greater degree. It can also be argued that because an interpretation of the SMFQ as measuring a unidimensional construct in this sample of 12-13 year olds and their parents can be supported, the unidimensionality would be strengthened at later time points, due to general maturation (Weiss & Garber, 2003). It has been suggested that prospective epidemiologic studies offer the best design for identifying early predictors of depression, and that early identification of predictors can aid interventions that seek to change the course of a disorder before it becomes chronic (Pine et al., 1999). The current study, although it uses data from only one point in time, is part of a larger prospective cohort study, and hopefully aids in providing proper preventive care for affected adolescents.

The current study examines the SMFQ in a multi-informant fashion, i.e. the factor structure and reliability is examined in both children and parents, and the parent-child agreement is studied in both parents of girls and girls, and parents of boys and boys. Having multiple informants allows for direct tests of comparability between different informants (Phillips, Lonigan, Driscoll, & Hooe, 2002), and deepens the understanding of both the SMFQ and the community in question. Comparisons between mothers and children are done combined, but also separated by sex. In this fashion, a better understanding of the ways mothers report depressive symptoms is obtained.

General limitations. As the SMFQ is based on criteria from the DSM-IV (APA, 2000), which views core symptoms of depression as similar in childhood, adolescence and adulthood, it can be questioned whether it is appropriate to apply adult criteria to children and adolescents. A developmental perspective might be more appropriate, in the sense that some depressive symptoms might vary with age and developmental level. The question of whether depression is expressed differently in childhood, adolescence and adulthood, and whether co-occurring disorders influence the developmental analyses of depression, remains to be answered (Weiss & Garber, 2003).

It can be questioned whether self-report on a questionnaire is a good measure of depressive symptomatology. It seems plausible that reporting bias in self-reports can result in minimisation or over-reporting of symptom severity, which would reduce the validity of the measure (Zimmerman, 2011). A reporting bias that is not random but systematic would make interpretations of questionnaires difficult.

In the context of the current study, there can be no assessments of convergent validity, because no parallel assessments of depression were administered in the TOPP study, in order to make the questionnaire as short as possible. Because there were no diagnostic interviews in the TOPP study, and the fact that an operationalisation of depression above a certain, essentially arbitrary, cut-off does not imply a clinical diagnosis of depression, psychometric assessments of criterion validity could not be undertaken in the current study.

Limitations in the sample. The current study only studies completers, i.e. only those mothers and children who in tandem completed the SMFQ were included in the analyses. It can be argued, however, that this has little impact on the analyses considering that the overall number of individuals of T5 in the TOPP-study was N =587 for mothers and N = 566 for children, whereas the current study has N = 505 for both mothers and children.

Validity concerns the actions one takes based on test scores and not the instrument itself (Messick, 1989). The current study finds that the SMFQ is a valid tool for studying depressive symptomatology within a population of adolescents in their early teens. Using the SMFQ in other age groups would necessitate an examination of the psychometric abilities of the SMFQ in these groups first. This could be done by examining the data from other time points in the TOPP-study. However, as a general rule, students are only allowed data from one time-point in the TOPP-study.

Attrition analyses indicate that the only predictor of non-participation when comparing T1 and T5 in the TOPP-study was low level of mothers' education (Karevold et al., 2009). In a

large review of studies the strongest predictors of mental health problems in children and adolescents were low household income and low parental education (Reiss, 2013). This could be a possible source of bias, which could mean that the prevalence of depressive symptoms and the relationship between SES indicators and depressive symptoms are underestimated.

Limitations in the measurement. A limitation of the SMFQ in the TOPP-study is that has only 12 items, whereas the original SMFQ has 13 items. The two measures are, however, highly correlated (r = .99; Nilsen et al., 2012).

Considering that the procedures of cross-culturally adapting instruments are not universally agreed upon (Gjersing, Caplehorn & Clausen, 2010), comparing results across different cultures, areas and languages might not be without its faults. It might be that concepts within an instrument, even if the translation process is considered valid, are unequal between languages and cultures (Gjersing et al., 2010). The MFQ, which the SMFQ is based on, was translated to Norwegian, back-translated and accepted by the original author (Sund et al., 2001). In the context of the current study, where the item regarding restlessness had high endorsement but low communalities in both girls and boys, questions regarding the appropriateness of the word "restless" ("rastløs") could be asked. One explanation for the low communalities could be that restlessness does not function as a symptom of depression in this sample, or it could be that the word itself is in inappropriate in a Norwegian context. A possible alternative to "rastløs" could be the word "urolig".

Conclusions & Implications

The current study examined the psychometric properties of the Short Mood and Feelings Questionnaire, and depressive symptomology in a community-based sample of 12-13 year old adolescents and their mothers. Through establishing structural validity, reliability and content validity, the current study attempted to validate the SMFQ. The SMFQ exhibited a unidimensional internal structure in all groups, thereby replicating findings across culture and language. Therefore, the use of SMFQ sum-scores as a continuous measure of severity of depressive symptoms can be supported. When comparing sum-scores between groups, children scored significantly higher than mothers, and girls scored significantly higher than boys. The ratio of girls to boys operationalised to be depressed was 4 to 3, and the parent-child agreement was moderate. Regarding SES indicators, boys who had single mothers had more depressive

symptoms, whereas girls living in families with lower income and an unsatisfactory economy had more depressive symptoms.

It was a goal of the current study to improve the knowledge about the SMFQ's psychometric abilities, because there are always demands for demonstrably valid instruments. A continued improvement in the identification of adolescents at risk of mental health problems, will contribute both to research and to society. In order to provide proper preventive care or implement appropriate interventions, one needs valid, reliable and thoroughly vetted tools for mapping and identifying at-risk or affected children and adolescents. These tools need to possess satisfactory psychometric abilities in order to be used as tools for identification, screening or evaluation. The current study has contributed in this regard. Further research into the data from the TOPP-study should seek to examine the psychometric properties of the SMFQ, and thereby validating it, for older age-groups.

References

Abela, J. R. Z., & Hankin, B. L. (2011). Rumination as a vulnerability factor to depression during the transition from early to middle adolescence: a multi-wave longitudinal study. *J. Abnorm. ChildPsychol. 120*, 259–271. doi:10.1037/a0022796

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological bulletin*, *101*(2), 213. doi:10.1037/0033-2909.101.2.213

Aggen, S. H., Neale, M. C., & Kendler, K. S. (2005). DSM criteria for major depression: Evaluating symptom patterns using latent-trait item response models. *Psychological Medicine, 35*, 475–487. doi:10.1017/S0033291704003563

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). American Psychiatric Association, Washington, DC.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text tev.). American Psychiatric Association, Washington, DC. doi:10.1176/appi.books.9780890423349

Angold, A. (2008). Dear potential MFQ user. *The MFQ.* Retrieved 19th of November 2014, from http://devepi.duhs.duke.edu/.%5Cinstruments%5CMFQ%20user.pdf

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40*, 57– 87. doi:10.1111/1469-7610 .00424

Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 5*(4), 237-249.

Angold, A., Costello, E. J., Pickles, E. J., Winder, F., & Silva, S. (1987). *The development of a questionnaire for use in epidemiological studies of depression in children and adolescents.* Medical Research Council, London, England.

Angold, A., Erkanli, A., Silberg, J., Eaves, L., & Costello, E. J. (2002). Depression scale scores in 8–17-year-olds: effects of age and gender. *Journal of child Psychology and Psychiatry*, *43*(8), 1052-1063. doi:10.1111/1469-7610.00232

Avenevoli, S., Stolar, M., Li, J., Dierker, L., & Merikangas, K. (2001). Comorbidity of depression in children and adolescents: Models and evidence from a prospective high-risk family study. *Biological Psychiatry, 49*, 1071–1081. doi:10.1016/S0006-3223(01)01142-8

Baglin, J. (2014). Improving Your Exploratory Factor Analysis for Ordinal Data: A Demonstration Using FACTOR. *Practical Assessment, Research & Evaluation*, *19*(5), 2.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin,* 107, 238–246. doi:10.1037/0033-2909.107.2.238

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the evaluation of personality scales. Journal of Personality, 54, 106-148. doi:10.1111/j.1467-6494.1986.tb00391.x

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, *24*(4), 445-455. doi:10.1207/s15327906mbr2404_4

Byrne, B. M. (2001). *Structural equation modeling with AMOS. Basic concepts, applications, and programming.* Mahwah, N.J: Lawrence Erlbaum Associates.

Carmine, E., & Zeller, R. (1979). *Reliability and validity assessment.* Newbury Park, CA: Sage.

Cerny, C.A., & Kaiser, H.F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research, 12*(1). doi:10.1207/s15327906mbr1201_3

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319. doi:10.1037/1040-3590.7.3.309

Cohen, J. (1988). Statistical power for the social sciences. *Hillsdale, NJ: Laurence Erlbaum and Associates*.

Cortina, J.M., (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology, 78*(1), 98–104. doi:10.1037/0021-9010.78.1.98

Costello, E. J., & Angold, A. (1988). Scales to assess child and adolescent depression: checklists, screens, and nets. *Journal of the American Academy of Child & Adolescent Psychiatry*, *27*(6), 726-737. doi:10.1097/00004583-198811000-00011

Costello, E. J., Benjamin, R., Angold, A., & Silver, D. (1991). Mood variability in adolescents: a study of depressed, nondepressed and comorbid patients. *Journal of Affective Disorders, 23*(4), 199-212. doi:10.1016/0165-0327(91)90101-w

Costello, A., Edelbrock, C., Kalas, R., Kessler, M., & Klaric, S. A. (1982). *The National Institute of Mental Health Diagnostic Interview Schedule for Children (DISC).* National Institute of Mental Health, Rockville, MD.

Costello, J. E., Erkanli, A. & Angold, A. (2006). Is there an epidemic of child or adolescent depression?. *Journal of Child Psychology and Psychiatry*, *47*(12), 1263-1271. doi:10.1111/j.1469-7610.2006.01682.x

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. doi:10.1007/bf02310555

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and psychological measurement*,*64*(3), 391-418. doi:10.1177/0013164404266386

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological bulletin*,*131*(4), 483. doi:10.1037/0033-2909.131.4.483

Derdikman-Eiron, R., Indredavik, M. S., Bakken, I.J., Bratberg, G.H., Hjemdal, O., & Colton, M. (2012). Gender differences in psychosocial functioning of adolescents with symptoms of anxiety and depression: longitudinal findings from the Nord-Trøndelag health study. *Soc. Psychiatry Psychiatr. Epidemiol. 47*, 1855–1863. doi:10.1007/s00127- 012-0492-y

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, *4*(3), 272. doi:10.1037/1082-989x.4.3.272

Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2007). Recurrence of major depression in adolescence and early adulthood, and later mental health, educational and economic outcomes. *The British Journal of Psychiatry*, *191*(4), 335-342. doi:10.1192/bjp.bp.107.036079

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods 9*, 466–491. doi:10.1037/1082-989X.9. 4.466

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, *16*(4), 625-641. doi:10.1080/10705510903203573

Friedman, E., & Anderson, I. (2014). *Handbook of depression* (2nd Ed.). Tarporley: Springer Healthcare. doi:10.1007/978-1-907673-79-5

Gjersing, L., Caplehorn, J. R., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC medical research methodology*, *10*(1), 13. doi:10.1186/1471-2288-10-13

Gustavson, K., von Soest, T., Karevold, E., & Røysamb, E. (2012). Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health, 12*(1), 918. doi:10.1186/1471-2458-12-918

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. *Journal of abnormal psychology*, *107*(1), 128. doi:10.1037/0021-843x.107.1.128

Heiervang, E., Stormark, K. M., Lundervold, A. J., Heimann, M., Goodman, R., Posserud, M. B., ... Gillberg, C. (2007). Psychiatric disorders in Norwegian 8-to 10-year-olds: An epidemiological survey of prevalence, risk factors, and service use. *Journal of the American Academy of Child and Adolescent Psychiatry*, *46*(4), 438 - 447. doi:10.1097/chi.0b013e31803062bf

Holte, A. (2012). Ti prinsipper for forebygging av psykiske lidelser. *Tidsskrift for Norsk Psykologforening*, 49 (7), 693-695.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 32*, 179-185. doi:10.1007/bf02289447

Horwath, E., Johnson, J., Klerman, G. L., & Weissman, M. M. (1992). Depressive symptoms as relative and attributable risk factors for first-onset major depression. *Archives of General Psychiatry*, *49*(10), 817-823. doi:10.1001/archpsyc.1992.01820100061011

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, *3*(4), 424. doi:10.1037/1082-989x.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. doi:10.1080/10705519909540118

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological assessment*, *15*(4), 446. doi:10.1037/1040-3590.15.4.446

John, O. P., & Benet-Martinez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.) *Handbook of Research Methods in Social and Personality Psychology* (pp. 339-369). New York: Cambridge University Press. doi:10.1017/cbo9780511996481.023

Karevold, E. (2008). *Emotional problems in childhood and adolescence: Predictors, pathways and underlying structure.* PhD Thesis. Oslo: Faculty of Social Sciences, University of Oslo.

Karevold, E., Coplan, R., Stoolmiller, M., & Mathiesen, K. S. (2011). A longitudinal study of the links between temperamental shyness, activity, and trajectories of internalising problems from infancy to middle childhood. *Australian Journal of Psychology, 63*(1), 36-43. doi: 10.1111/j.1742-9536.2011.00005.x

Karevold, E., Røysamb, E., Ystrom, E., & Mathiesen, K. S. (2009). Predictors and pathways from infancy to symptoms of anxiety and depression in early adolescence. *Developmental psychology*, *45*(4), 1051. doi:10.1037/a0016123

Karver, M. S. (2006). Determinants of multiple informant agreement on child and adolescent behavior. *Journal of abnormal child psychology*, *34*(2), 242-253. doi:10.1007/s10802-005-9015-6

Kazdin, A. E. (1989). Identifying depression in children: A comparison of alternative selection criteria. Journal of Abnormal Child Psychology, 17, 437– 455. doi:10.1007/bf00915037

Kent, L., Vostanis, P., & Feehan, C. (1997). Detection of major and minor depression in children and adolescents: evaluation of the Mood and Feelings Questionnaire. *Journal of Child Psychology and Psychiatry*, *38*(5), 565-573. doi:10.1111/j.1469-7610.1997.tb01543.x

Kjeldsen, A. (2013). *Longitudinal profiles of externalizing behavior problems from infancy to mid-adolescence: early predictors, timing of risk factors, and late adolescence outcomes.* Doctoral thesis. University of Oslo, Oslo, Norway

Kovacs, M. (1983). *The Children's Depression Inventory: A self-rated depression scale for school-aged youngsters.* Unpublished manuscript, University of Pittsburgh School of Medicine.

Loevinger, J. (1957). Objective tests as instruments of psychological theory: Monograph Supplement 9. *Psychological reports*, *3*(3), 635-694. doi:10.2466/pr0.1957.3.3.635

Lorenzo-Seva, U., & Ferrando, P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers, 38*(1), 88-91. doi:10.3758/bf03192753

Lundervold, A. J., Breivik, K., Posserud, M. B., Stormark, K. M., & Hysing, M. (2013). Symptoms of depression as reported by Norwegian adolescents on the Short Mood and Feelings Questionnaire. *Frontiers in psychology*, *4*. doi:10.3389/fpsyg.2013.00613

May, R. (1989). *Love and Will* (New Delta ed., p. 243). New York: Dell.

Mathiesen, K., Sanson, A., Stoolmiller, M., & Karevold, E. (2008). The Nature and Predictors of Undercontrolled and Internalizing Problem Trajectories Across Early Childhood. *Journal of Abnormal Child Psychology, 37*(2), 209-222. doi: 10.1007/s10802-008-9268-y

McIntire, S. A., & Miller, L. A. (2007). *Foundations of psychological testing: A practical approach*. Sage Publishing.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical care*, *44*(11), 69-77. doi:10.1097/01.mlr.0000245438.73837.89

Messer, S. C., Angold, A., Costello, E. J., & Loeber, R. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents: Factor composition and structure across development. *International Journal of Methods in Psychiatric Research*, 5(4), 251-262.

Messer, S. C., & Gross, A. M. (1994). Childhood depression and aggression: A covariance structure analysis. *Behaviour research and therapy*, *32*(6), 663-678. doi:10.1016/0005-7967(94)90023-x

Messick, S. (1989). Validity. In Linn, R. L. (Ed.), *Educational Measurement* (13–103). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. doi:10.1037/0003-066X.50.9.741

Muchinsky, P. M. (1996). The Correction for Attenuation. *Educational and Psychological Measurement, 56*(1), 63–75. doi:10.1177/0013164496056001004

Muthén, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171–189. doi:10.1111/j.2044-8317.1985.tb00832.x

Muthen, L., & Muthen, B. (1998- 2012). *Mplus Users Guide (7th Ed.).* Los Angeles, CA: Muthen & Muthen.

Mykletun A, Knudsen, A. K., & Mathiesen, K.S. (2009). *Rapport 2009:8. Psykiske lidelser i Norge: Et folkehelseperspektiv.* Oslo: Nasjonalt folkehelseinstitutt. Retrieved 19th of January, 2015, from http://www.fhi.no/dav/68675aa178.pdf

Nilsen, W., Karevold, E., Røysamb, E., Gustavson, K., & Mathiesen, K. S. (2012). Social skills and depressive symptoms across adolescence: Social support as a mediator in girls versus boys. *Journal of Adolescence*. doi:10.1016/j.adolescence.2012.08.005

Norwegian Institute of Public Health (2015). *TOPP-studien.* Retrieved 25th of March, 2015 from http://www.fhi.no/artikler/?id=66188

Oldehinkel, A. J., Wittchen, H. U., & Schuster, P. (1999). Prevalence, 20-month incidence and outcome of unipolar depressive disorders in a community sample of adolescents. *Psychological Medicine*, *29*(03), 655-668. doi:10.1017/s0033291799008454

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460. doi:10.1007/bf02296207

Olsson, G. I., von Knorring, A. L. (1999) Adolescent depression: prevalence in Swedish high-school students. *Acta Psychiatr. Scand., 99*(5), 324-331. doi:10.1111/j.1600-0447.1999.tb07237.x

Pallant, J. (2013). *SPSS Survival Manual. A step by step guide to data analysis using IBM SPSS.* Berkshire, England: Open University Press.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74,* 525-556. doi:10.3102/00346543074004525

Phillips, B. M., Lonigan, C. J., Driscoll, K., & Hooe, E. S. (2002). Positive and negative affectivity in children: A multitrait-multimethod investigation. *Journal of Clinical Child and Adolescent Psychology*, *31*(4), 465-479. doi:10.1207/s15374424jccp3104_6

Pine, D. S. Cohen, E. Cohen , P. Brook , J. (1999). Adolescent depressive symptoms as predictors of adult depression: Moodiness or mood disorder? *American Journal of Psychiatry, 156*, 133-135. doi:10.1176/ajp.156.1.133

Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Social Science & Medicine, 90*, 24–31. doi:10.1016/j.socscimed.2013.04.026

Renouf, A. G., & Kovacs, M. (1994). Concordance between mothers' reports and children's self-reports of depressive symptoms: a longitudinal study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *33*(2), 208-216. http://dx.doi.org/10.1097/00004583-199402000-00008

Richter, J. & Sund, A. M. (2013). Måleegenskaper ved den norske versjonen av Mood and Feelings Questionnaire - Humøret Ditt (MFQ). *PsykTestBarn, 1:1.*

Rigdon, E. E., & Ferguson Jr, C. E. (1991). The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research 28*(4), 491-497. doi:10.2307/3172790

Sadock, V. A., Sadock, B. J., Kaplan, H. I. (2003). *Kaplan & Sadock's synopsis of psychiatry: behavioral sciences/clinical psychiatry*. Philadelphia: Lippincott Williams & Wilkins.

Sanne, B., Dahl, A., & Tell, G. (2001). Depresjon – samfunnsøkonomiske perspektiver. *Tidsskrift for Den norske legeforening.* Retrieved 19th of January, 2015, from http://tidsskriftet.no/article/270495/

Sharp, C., Goodyer, I. M., & Croudace, T. J. (2006). The Short Mood and Feelings Questionnaire (SMFQ): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *Journal of abnormal child psychology*, *34*(3), 365-377. doi:10.1007/s10802-006-9027-x

Sivertsen, B., Harvey, A. G., Lundervold, A. J., & Hysing, M. (2014). Sleep problems and depression in adolescence: results from a large population-based study of Norwegian adolescents aged 16–18 years. *European Child & Adolescent Psychiatry*, *23*(8), 681-689. doi:10.1007/s00787-013-0502-y

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 1904-1920*, *3*(3), 271-295. doi:10.1111/j.2044-8295.1910.tb00206.x

Steiger, J., & Lind, J. (1980). Statistically based tests for the number of common factors. *Annual Meeting of the Psychometric Society,* Iowa City, IA.

Steinhausen, H. C., Winkler, M. (2003) Prevalence of affective disorders in children and adolescents: findings from the Zurich Epidemiological Studies. *Acta Psych Scand 108*(suppl 418), 20-23. doi:10.1034/j.1600-0447.108.s418.5.x

Streiner, D., & Norman, G. (1989). *Health Measurement Scales: A practical guide to their development and use*. New York, NY: Oxford University Press.

Sund, A. M. (2004). *Development of depressive symptoms in early adolescence: the youth and mental health study* (Doctoral thesis). Trondheim: The Faculty.

Sund, A. M, Bjelland I., Holgersen H., Israel P. & Plessen K. J. (2012). Om depresjon hos barn og unge med vekt på biologiske modeller. *Tidsskrift for Norsk Psykologforening, 49*, 30-39.

Sund, A. M., Larsson, B., & Wichstrøm, L. (2001). Depressive symptoms among young Norwegian adolescents as measured by the Mood and Feeling Questionnaire (MFQ). *European Child and Adolescent Psychiatry, 10*, 222–229. doi:10.1007/s007870170011

Sund, A. M., Larsson, B., & Wichstrom, L. (2011). Prevalence and characteristics of depressive disorders in early adolescents in central Norway. *Child and adolescent psychiatry and mental health, 5*(1), 28. doi:10.1186/1753-2000-5-28

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2*(53). doi:10.5116/ijme.4dfb.8dfd

Thapar, A., Collishaw, S., Pine, D. S., & Thapar, A. K. (2012). Depression in adolescence. *The Lancet*, *379*(9820), 1056-1067. 1056–1067. doi:10.1016/s0140-6736(11)60871-4

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–20. doi:10.1037/a0023353

Trochim, W. M. K. (2006). Types of Reliability. *Social Research Methods.* Retrieved from http://www.socialresearchmethods.net/kb/reltypes.php

Tucker, L. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 38, 1–10. doi:10.1007/BF02291170

Turner, N., Joinson, C., Peters, T. J., Wiles, N., & Lewis, G. (2014). Validity of the Short Mood and Feelings Questionnaire in late adolescence. *Psychological Assessment, 26*(3), 752-762. doi:10.1037/a0036572

Watkins, E., & Moulds, M. (2005). Distinct modes of ruminative self-focus: impact of abstract versus concrete rumination on problem solving in depression. *Emotion 5*, 319–328. doi:10.1037/1528-3542.5. 3.319

Weiss, B., & Garber, J. (2003). Developmental differences in the phenomenology of depression. *Development and Psychopathology*, *15*(02), 403-430. doi:10.1017/s0954579403000221

Wichstrøm, L., (1999). The emergence of gender difference in depressed mood during adolescence: The role of intensified gender socialization. *Develomental Psychology, 35*, p. 232–45. doi:10.1037/0012-1649.35.1.232

World Health Organization (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* Geneva: World Health Organization.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, *12*(3), 1-26.

Yu, C. Y. (2002). Evaluating cut-off criteria of model fit indices for latent variable models with binary and continuous outcomes. Master's thesis, University of California, Los Angeles, CA.

Zimmerman, M. (2011). Tools for Depression: Standardized Rating Scales. Retrieved 20th of January, 2015, from http://www.medscape.org/viewarticle/749921

Appendices

Appendix A

Core symptoms of depression in DSM-IV and ICD-10 (APA, 2000; WHO, 1992).

DSM-IV	*ICD-10*
Depressed or irritable mood[+]	Depressed mood*
Diminished interest or pleasure in activities[+]	Loss of interest and enjoyment*
Fatigue or loss of energy	Increased fatiguability; diminished activity*
Change in appetite; weight change	Diminished appetite
Insomnia or sleeping too much	Disturbed sleep
Feelings of worthlessness or guilt	Ideas of guilt and unworthiness
Diminished ability to think or concentrate	Reduced concentration and attention
Recurrent thoughts of death	Thoughts or acts of self-harm or suicide
Agitation or psychomotor retardation	-
-	Reduced self-esteem and self-confidence
-	Bleak and pessimistic views of the future
+: At least 1 of 2 must be present for two weeks. At least 5 of 9 symptoms need to be present nearly every day.	**: At least 2 of 3 must be present for two weeks. 4, 6 and 8 symptoms indicate mild, moderate and severe, respectively.*

Appendix B

Items in the SMFQ	
English	*Norwegian*
I felt miserable or unhappy	Jeg var lei meg eller ulykkelig
I didn't enjoy anything at all	Jeg var ikke glad for noe
I felt so tired I just sat around and did nothing	Jeg følte meg så trøtt at jeg bare ble sittende uten å gjøre noen ting
I was very restless	Jeg var veldig rastløs
I felt I was no good any more	Jeg følte meg lite verdt
I cried a lot	Jeg gråt mye
I found it hard to think properly or concentrate	Jeg syntes det var vanskelig å tenke klart eller å konsentrere seg
I hated myself	Jeg hatet meg selv
I was a bad person	Jeg følte meg som et dårlig menneske.
I felt lonely	Jeg følte meg ensom
I thought nobody really loved me	Jeg tenkte at ingen egentlig var glad i meg
I thought I could never be as good as other kids	Jeg tenkte at jeg aldri kunne bli så god som andre barn
I did everything wrong	Jeg gjorde alt galt

Appendix C

Humøret Ditt (kortversjon) / SMFQ.			
Her er en liste over forskjellige plagsomme følelser og tanker man av og til kan ha. Tenk på **de to siste ukene** og kryss av for hvor ofte du har følt eller tenkt noe av det som står nedenfor: (sett ett kryss på hver linje)	1 Stemmer ofte	2 Stemmer noen ganger	3 Stemmer sjelden
Jeg var lei meg eller ulykkelig			
Jeg følte meg så trøtt at jeg bare ble sittende uten å gjøre noen ting			
Jeg var veldig rastløs			
Jeg var ikke glad for noe			
Jeg følte meg lite verdt			
Jeg gråt mye			
Jeg hatet meg selv			
Jeg tenkte at jeg aldri kunne bli så god som andre barn			
Jeg følte meg ensom			
Jeg tenkte at ingen egentlig var glad i meg			
Jeg følte meg som et dårlig menneske			
Jeg syntes jeg gjorde alt galt			

Appendix D

Humøret Ditt (kortversjon) / SMFQ.			
Mange kan være nedfor fra tid til annen, og noen er plaget av triste tanker. Tenk på **de siste to ukene** og angi hvor godt hvert av utsagnene stemmer på barnet/ungdommen din (sett ett kryss på hver linje). **Han/hun:**	1 Stemmer ofte	2 Stemmer noen ganger	3 Stemmer sjelden
var lei seg eller ulykkelig			
følte seg så trøtt at han/hun bare ble sittende uten å gjøre noen ting			
var veldig rastløs			
var ikke glad for noe			
følte seg lite verdt			
gråt mye			
hatet seg selv			
tenkte at han/hun aldri kunne bli så god som andre ungdommer			
følte seg ensom			
tenkte at ingen egentlig var glad i ham/henne			
følte seg som et dårlig menneske			
syntes han/hun gjorde alt galt			

Appendix E

Appendix E – Indicators of socioeconomic status for mothers of girls and girls' mean SMFQ scores.

Type	*Level*	*%age*	*N*	*SMFQ M score*	*Type*	*Level*	*%age*	*N*	*SMFQ M score*
Mother's education	Elementary school	4.5%	12	3.8	Partner's education	Elementary school	7.3%	16	4.4
	1-2 years high school	19.5%	52	4.4		1-2 years high school	23.3%	51	3.8
	3 years high school	22.5%	60	3.8		3 years high school	13.2%	29	3.6
	1-4 years college or university	25.8%	69	4.1		1-4 years college or university	21%	46	4.3
	>4 years college or university	27.7%	74	3.4		>4 years college or university	35.2%	77	3.1
Mother's employment status	No gainful employment	13.4%	36	4.8	Partner's employment status	No partner	5.6%	13	4.2
	<50% part-time work	6.3%	17	5.4		No gainful employment	1.7%	4	3
	50-80% part-time work	26.1%	70	3.7		<50% part-time work	0.9%	2	0*
	80-100% full-time work	54.1%	145	3.6		50-80% part-time work	2.2%	5	1.8
						80-100% full-time work	89.7%	208	3.8*
Family economy coping	Very poor	0.4%	1	11^{+}	Family income in NOK	<100k	0.7%	2	7.5^{+}
	Poor	2.6%	7	2.3^{+}		100k – 150k	1.9%	5	6.8^{+}
	OK	22.3%	60	5^{+}		150k – 200k	4.8%	13	5.2^{+}
	Well	44.2%	119	4^{+}		200k – 300k	9.7%	26	2.7^{+}
	Extremely well	30.5%	82	2.9^{+}		300k – 400k	11.9%	32	5.5^{+}
						400k – 500k	27.5%	74	4.0^{+}
						>600k	43.5%	117	3.3^{+}

* denotes $p < .01$; + denotes significant main effect.

Appendix F

Appendix F – Indicators of socioeconomic status for mothers of boys and boys' mean SMFQ scores.

Type	*Level*	*%age*	*N*	*SMFQ M score*	*Type*	*Level*	*%age*	*N*	*SMFQ M score*
Mother's education	Elementary school	5.8%	13	1.7	Partner's education	Elementary school	4.6%	9	3.8
	1-2 years high school	16.0%	36	3.6		1-2 years high school	19.5%	38	3
	3 years high school	26.2%	59	3.1		3 years high school	14.9%	29	2.4
	1-4 years college or university	24.4%	55	2.7		1-4 years college or university	23.1%	45	3.5
	>4 years college or university	27.6%	62	3.2		>4 years college or university	37.9%	74	2.4
Mother's employment status	No gainful employment	13.8%	31	2.2	Partner's employment status	No partner	7.1%	15	5.1
	<50% part-time work	6.7%	15	2.8		No gainful employment	2.8%	6	4.7
	15-80% part-time work	23.1%	52	3.6		<50% part-time work	0.5%	1	0
	80-100% full-time work	56.4%	127	3.0		15-80% part-time work	0.9%	2	0.5
						80-100% full-time work	88.7%	188	2.8
Family economy coping	Very poor	0%	0	-	Family income in NOK	<100k	0%	0	-
	Poor	2.2%	5	4		100k – 150k	0.9%	2	2.5
	OK	19.1%	43	3.6		150k – 200k	3.6%	8	3.4
	Well	49.8%	112	3.0		200k – 300k	9%	20	3.6
	Extremely well	28.9%	65	2.6		300k – 400k	12.7%	28	3.4
						400k – 500k	23.5%	52	3.3
						>600k	50.2%	111	2.8